PROJECT ABSTRACT

Music Genre Classification using Machine Learning

Topic ID: K-1220 | Branch: AI & Machine Learning | Type: Software | Availability: Built-to-order

01. Title

Music Genre Classification using Machine Learning — 10-genre classification from MFCC and spectral features.

02. Introduction / Background

Streaming services classify every uploaded track by genre, mood and instrumentation; the academic ancestor of that pipeline is GTZAN — 1,000 thirty-second clips across 10 genres, published by Tzanetakis & Cook in 2002 and cited by hundreds of papers since.

Raw audio is useless to a classifier directly: the work is feature extraction. MFCCs mimic human hearing to capture timbre, chromagrams capture harmony, spectral contrast captures texture, and tempo captures groove — aggregated into a 58-dimensional fingerprint per clip. This project builds that full pipeline, trains a random-forest classifier, and delivers a track-classifier demo with mel-spectrogram fingerprints and per-genre error analysis.

03. Problem Statement

Sort 30-second music clips into 10 genres from audio features alone, with honest stratified evaluation that prevents the model from memorizing artists. The pedagogical goal is the full audio-ML pipeline — framing, spectral features, aggregation, classification — and the analytical goal is the per-genre error story: which genres separate cleanly and which (rock/metal, disco/pop) genuinely confuse, matching the published literature.

04. Objectives

- Build a librosa feature pipeline: 13 MFCCs + deltas, 12-bin chroma, spectral centroid/rolloff/contrast, ZCR, tempo → 58 features per clip.
- Train a 300-tree random forest on standardized fingerprints (design target: 84.2% accuracy).
- Evaluate on a stratified 80/20 holdout with per-genre precision/recall/F1 and confusion matrix.
- Build a track-classifier demo: spectrogram fingerprint, predicted genre, top-5 probability bars, genre explorer.
- Analyze genre confusions against the literature (rock–metal, disco–pop).
- Document audio features, MFCC theory and evaluation discipline for the viva.

05. Proposed Methodology

- Load the GTZAN collection (1,000 clips, 10 × 100, 30 s, 22,050 Hz mono); verify integrity.
- Frame each clip into overlapping windows; compute per-frame MFCCs, chroma, spectral features, ZCR.
- Aggregate frame features (mean + variance) into the 58-dimensional clip fingerprint; standardize.
- Stratified 80/20 split, artist-aware where possible; train the random forest.
- Evaluate the holdout once: accuracy, per-genre F1, confusion matrix.
- Serve the classifier in the web demo with spectrogram fingerprints and probability rendering.

06. System Architecture / Working

Audio clips → framing → librosa feature extraction → mean/variance aggregation → standardization → random forest → genre probabilities → demo UI. The demo renders a mel-spectrogram-style fingerprint per track (12 mel bands × time) alongside the probability bars so users see what the model 'hears'.

Dataset: GTZAN Genre Collection (Tzanetakis & Cook, 2002) — 1,000 clips, 10 genres. Task: 10-class audio classification. Model: random forest, 300 trees. Metrics: accuracy, per-genre F1 — design targets, never claimed as measured.

07. Hardware / Software Requirements

- Hardware: standard laptop/desktop, 8 GB RAM; CPU feature extraction and training in minutes.
- Software: Python 3.10+, librosa, scikit-learn, NumPy, pandas, Matplotlib, Jupyter.
- Demo: any modern browser; single-file web app runs offline after download.

08. Expected Outcomes

- librosa feature-extraction pipeline producing 58 features per clip.
- Trained random-forest genre classifier with inference code.
- Training notebook with stratified evaluation and confusion analysis (84.2% design target).
- Track-classifier + genre-explorer demo with spectrogram fingerprints.
- Report, PPT and viva Q&A; on MFCCs, spectrograms, forests and artist-aware splits.

09. Applications

- Music library organization: automatic genre tagging of personal collections.
- Recommendation cold-start: genre profiles for new catalog tracks.
- DJ software: key/genre-aware set planning aids.
- Catalog analytics: genre distribution and trend studies for labels.
- Music education: interactive demonstrations of timbre vs harmony vs rhythm.

10. Future Scope

- CNN on mel-spectrograms as the deep-learning extension.
- Multi-label tagging: mood, instruments and era alongside genre.
- Larger datasets (Free Music Archive) for data-hungry models.
- Real-time track identification in the Shazam style.
- Audio transformers as a state-of-the-art comparison.